Journal of Computer Science

Research Article

Refining Chest X-ray Interpretation with Deep Transfer Learning Techniques

R. Sriramkumar[1], K. Selvakumar[1] and J. Jegan[2]

[1]*Department of Information Technology, Annamalai University, Chidambaram, Tamil Nadu, India*
[2]*Department of Computer Science and Engineering, SRM Institute of Science and Technology, Tiruchirappalli, Tamil Nadu, India*

Article history
Received: 01-05-2025
Revised: 19-06-2025
Accepted: 23-06-2025

Abstract: Chest X-rays are essential diagnostic tools for thoracic diseases, but different doctors' interpretations can differ greatly, which frequently results in inconsistent diagnoses. By using deep transfer learning techniques, this study seeks to improve the accuracy of Chest X-ray interpretations. More accurate and effective analysis can be accomplished with the growing potential of artificial intelligence (AI) in medical imaging, especially through convolutional neural networks (CNNs). It takes a lot of resources to train these models on sizeable annotated datasets, though. By optimising models that have already been trained on general datasets for particular medical imaging tasks, transfer learning provides a solution. In order to improve image quality, this study presents a dual-model framework that makes use of MobileNetV2 and InceptionV3. It is optimised using sophisticated preprocessing techniques like Contrast Limited Adaptive Histogram Equalisation (CLAHE) and white balance correction. Together with these improvements, data augmentation fills in the existing gaps in deployable, lightweight models for real-time applications in clinical settings with limited resources. When tested on illnesses like lung cancer, pneumonia, and tuberculosis, the system demonstrated notable gains in sensitivity and classification accuracy when compared to conventional diagnostic techniques. Additionally, the models show promise for being incorporated into clinical workflows, which would help radiologists detect diseases early and cut down on diagnostic delays. All things considered, this strategy helps provide healthcare in a more reliable, effective, and easily accessible manner.

Keywords: Ensemble, MobileNet, Deep Learning, Chest X-ray, CNN, Inception V3, Healthcare

Introduction

Medical image analysis systems are now much more capable thanks to recent advancements in artificial intelligence, especially in deep learning (LeCun *et al.*, 2015). Because they can automatically extract hierarchical features from image data, Convolutional Neural Networks (CNNs) have become a potent tool in this field, showing promise in tasks like segmentation, detection, and classification (Litjens *et al.*, 2017). Despite achieving high classification accuracy in Chest X-ray images, architectures such as ResNet and VGG are less suited for use in real-time or mobile clinical applications due to their high computational requirements (Rajpurkar *et al.*, 2017).

In order to get around these restrictions, this study suggests a dual-model framework that combines InceptionV3, a deeper network that can capture multi-scale image features, with MobileNetV2, which is renowned for its lightweight and effective design. Lakshmanan *et al.* (2024a) reviewed the wider perspective of how AI is changing the workforce and the labor markets. Their results highlight how automation technologies may transform industries, such as healthcare, through the introduction of AI-based decision systems. That is consistent with our work, in which deep transfer learning models powered by AI are used in medical imaging, and here, it has both efficiency and accuracy advantages (Hole *et al.*, 2024). A framework using deep neural networks was introduced for identifying tuberculosis in chest radiographs. The approach delivered strong accuracy and clear diagnostic outcomes, supporting its potential use in medical decision-making. This direction connects with our study,

which applies deep learning for chest disease detection to strengthen reliability in healthcare (Maheswari *et al.*, 2024). For the diagnosis of pulmonary diseases such as pneumonia, tuberculosis, and lung cancer, chest radiography is still an essential diagnostic technique (World Health Organization, 2022). Reliable computer-aided diagnosis (CAD) systems must be developed because manual X-ray interpretation can be laborious and subject to inter-observer variability (Ippolito *et al.*, 2023). Transfer learning is used to adapt pretrained models to the target medical imaging domain, which speeds up model convergence and eliminates the need for large labeled datasets (Shin *et al.*, 2016).

Among the oldest types of imaging tools to diagnose a lung pathology like lung cancer, tuberculosis and pneumonia are the chest X-rays. Chest X-rays form a foundation diagnostic method, yet, as with any Chest X-ray interpreter, it continues to be an extremely demanding and controversial method in professional circles. Delays or failure lead to a disastrous early-stage presentation that has dire effects on disease diagnosis and is highly destructive to the control of the interplay between a multi-faceted care pathway. More advanced computer systems are required that can assist in the analysis of Chest X-rays to the advantage and support of the radiologists. With the high level of intelligence, the systems would automate the complex tasks (Thimoteo *et al.*, 2022).

The deployment of transfer learning techniques within deep neural networks is without a doubt the fastest maturing area. You can see this in the fact that you can train new advanced models with restricted training data for logically built ‘simple cases’ that are meant to be more complex.

Transfer learning helps improve the effectiveness and efficiency of specific task models, such as disease detection from medical images, by further leveraging existing frameworks trained on large datasets. Two popular MobileNet and Inception V3 CNN architecture performances were compared in terms of disease prediction based on Chest X-ray images. MobileNet is known for its low computational requirements and lightweight, which enables its use in resource-limited environments like point-of-care devices and mobile phones. This is due to the use of depthwise separable convolutions which lowers the parameter count, computational expense, and increases accuracy. Inception V3 is deeper and more sophisticated than the former; it features multiple convolutional filters of various sizes to capture different scale features at once. The architecture has been used effectively in a number of image classification competitions for classification purposes (Saritha *et al.*, 2022).

The models MobileNet and Inception V3 will be fine-tuned for this activity with a dataset containing labeled Chest X-ray images. The models were modified for enhancement of pathological pattern detection in Chest X-rays for the diagnosis of diseases (Huang *et al.*, 2022). The results obtained after the evaluation were assessed based on Area Under the Receiver Operating Characteristic Curve. Accuracy, sensitivity, specificity, and the Area Under the Receiver Operating Characteristic Curve is just one of the statistics castoffs to evaluate the accuracy of these models. In addition, the results can be used in the future to create AI derived diagnostic measures that can potentially be incorporated into the mainstream medical process to provide effective patient management and better clinical care and outcomes (Gaur *et al.*, 2023).

Research Objective

This study introduces a deep transfer learning model to analyse Chest X-rays to detect thoracic diseases and achieve high detection rates within a short period. X-rays can be interpreted manually, which is often tedious and needs sophisticated skills, and not all healthcare professionals can interpret the X-rays the same, causing a discrepancy in the results. To address these weaknesses, the proposed system modulates the pre-trained CNN architectures on radiologist-annotated datasets and uses the latest image preprocessing methods, such as CLAHE and white balance balancing, to enhance the image. The data augmentation techniques are rotation, scaling, flipping or reversing to enhance the variety of a dataset, generalisation of the model and overfitting. It is based on rigorous validation procedures, including cross-validation and holdout tests, and its magnitude is quantified in terms of accuracy, sensitivity, specificity and F1-score. The framework is useful for identifying important conditions, such as pneumonia, tuberculosis, and lung cancer. Its small size renders it practical in real-time application implementation as well as in mobile medical care environments and facilitates convenient clinical implementation via an easy integration with hospital information systems. On the whole, the model suggested will make the diagnostic process more accurate, decrease waiting periods, help radiologists make better decisions, and improve patient outcomes in diseases of the chest.

Literature Survey

The study demonstrates an improvement in the accuracy of detection of disease in Chest X-ray images, specifically COVID-19 with pretrained CNNs. The methods used to detect COVID-19 utilize pretrained CNNs based on transfer learning. They compare the performance of different models concerning diagnosis and the preprocessing steps used. It is shown that transfer learning has a great impact on the accuracy of COVID-19 diagnosis. The study assesses the effectiveness of the proposed model in relation to the other models. The study shows that using transfer learning enhances the accuracy level of COVID-19 diagnosis significantly. The study makes use of pretrained CNNs and transfer learning techniques for the identification of COVID-19.

Various model choices, diagnostic performance, and preprocessing are analyzed. The results of the performed research is juxtaposed with the conclusions of the developed approach evaluating its effectiveness with existing methods (Singh *et al.*, 2024).

In methodological terms, the aspect in this study is to increase the accuracy of Chest X-ray scan diagnosis of lung diseases using deep transfer learning technology. The accuracy of 96.21% has been achieved. Hybrid CNNs, transformers and Big Transfer Explainable AI (XAI) are regarded as credible and clear models of deep learning (Ifty *et al.*, 2024). The Xception model that was built performed the best out of all participants with an accuracy of 96.21%. Other ensembles and models were between 86% and 93% in accuracy. Compared to SCAXN, NIWRHSO, and SBIGRU approaches, the suggested deep learning model with pre-trained transfer learning performed better than the other models in multiclass classification of diseases using Chest X-ray images. An XGBoost classifier is used to classify tuberculosis, which has been extracted with DenseNet201 (Boyina *et al.*, 2024) With the SCAXN, NIWRHSO and SBIGRU techniques, the proposed model employing deep learning Using Pre-trained transfer learning outperformed the current techniques on accuracy of multiclass illness identification in Chest X-ray images. Section SMOTE, SCAXN, NIWRHSO, SBIGRU are the illness recognition and classification methods. Krishnamoorthy *et al.* (2024) proposed a deep learning approach to multiclass lung diseases diagnoses in CXR images and it yielded the best results compared to all other methods tested (Maquen-Niño *et al.*, 2024).

The highest projection accuracy for pneumonia in Chest X-ray images with transfer learning was noted as 0.91% using ResNet50, whereas the remaining models, DenseNet and VGG19, achieved accuracies of 0.87% and 0.86%. Its convergence with the CNN model resulted in an overall classification accuracy of 83.57% with ResNet50. The Custom CNN model did not provide improvements to the accuracy, resulting in an achievement of 78.25%, compared to the pretrained architectures which outperformed ResNet at an accuracy of 83.57%. Custom CNN was trained with a weighted loss methodology alongside pre-trained frameworks. Enhancements in the accuracy of the diagnostic were primary motivated objectives alongside limited overfitting, which puts great focus on the necessity of regularization approaches. The figure of merit on the modified CNN+VGG19 model regarded with VGG19 for disease recognition in Chest X-ray images indicates such evaluation is most informative when done pre and post alterations to the model. Overfitting was an additional factor that led to reduced metric values on the deep learning architecture with CNN+VGG19 (Lakshmanan *et al.*, 2022; Zanaj *et al.*, 2024; Kumar *et al.*, 2023).

Studies utilising ResNet50, DenseNet, and VGG for Chest X-ray classification are included in the literature review. Although these models are accurate, few of them employ real-time, lightweight architectures. Additionally, the current approaches employ little preprocessing and do not provide a comprehensive evaluation of class-wise metrics. To close these gaps, we use 5-fold cross-validation, white balance correction, and CLAHE to increase diagnostic accuracy and generalizability (Hussain *et al.*, 2024).

This work further uses deep learning transfer models, specifically VGG16, for juvenile pneumonia diagnosis from Chest X-ray images. They are based on custom CNNs, as well as transfer learning with VGG16, Inception v3, and ResNet 152 v2. The model was able to get 97.18% recall and 92.63% accuracy. It assists in the enhancement of X-ray screening of juvenile pneumonia (Ouerhani *et al.*, 2023). The journal paper discusses the application of machine learning (ML) and deep learning (DL) for the early and precise detection of skin cancers, particularly melanoma. The authors note that skin cancer incidence is on the rise and automated, accurate diagnostic tools are urgently needed. Support vector machines (SVM), k-nearest neighbors (KNN), and random forest (RF) have been used to classify lesions based on features extracted. The paper, however, points out that deep learning, in particular, convolutional neural networks (CNNs) outperform other methods because they can learn to detect features from the image without needing preprocessing for the data. The study aims at classifying dermoscopic images using several CNN architectures and transfer learning models including VGG16 and Resnet50.

To improve model accuracy, various data preprocessing techniques such as image augmentation and normalization are applied. The authors performed experiments on publicly available datasets, such as ISIC, and were able to attain elevated levels of accuracy, sensitivity, and specificity in the classification of skin lesions. Moreover, they highlight issues such as data imbalance, model overfitting, and explainable AI in the context of clinical applicability. This paper claims DL-based techniques have great potential in aiding dermatologists in the diagnosis of skin cancer by offering quicker and more dependable results. Incorporating additional datasets, real-time clinical workflows, and enhancing generalizability of the models are among the plans for future research (Mazhar *et al.*, 2023).

A blockchain-based framework was proposed to improve traceability and safety in medical waste management. The work demonstrated how decentralized systems can enhance transparency, security, and environmental protection in healthcare operations (Lakshmanan *et al.*, 2025).

With regard to medical imaging, deep learning has remarkably expanded processes how complex features are extracted from large data sets through automation. Moreover, CNNs, or convolutional neural networks, have showcased their capabilities in image classification,

segmentation, and detection for MRI, CT, and X-rays. These approaches shift the burden of diagnosis from human experts and handcrafted features to automated systems, which subsequently enhances accuracy and efficiency. Deep learning is capable of hierarchical representation learning, and therefore, it is trained to grow that detects subtle patterns which are often associated or indicate diseases. From brain disorders and cancer detection to cardiac imaging, the spectrum of applications is immense. A few challenges still exist such as the availability of data, data annotation, clinical integration, interpretability, and most importantly, the ease of use in cinemas. Regardless, data lacking challenges are mitigated with the use of transfer learning and data augmentation. In addition, explainable AI systems are in development to build trust and transparency and enhance the understanding to improve the overarching system. All in all, efficient deep learning is expected to enhance medical imaging that enables early diagnosis and optimized treatment tailored for individual needs (Kumar Singh *et al.*, 2021).

This paper focuses on the application of COVID-19 infection detection using deep learning techniques and Chest X-ray images. Convolutional Neural Networks (CNNs) are employed for image classification and feature extraction because of their effectiveness. The authors sharpen the diagnostic accuracy using transfer learning with pre-trained models of VGG16, ResNet50, and InceptionV3, even with small datasets. The study places considerable attention on the underrepresentation of certain classes, specifically image normalization and augmentation. Evaluation is conducted using a set of metrics comprising accuracy, sensitivity, specificity, and F1-score, among others. Findings indicate that patients with confirmed COVID-19 diagnoses can be efficiently and accurately automated distinguishing them from patients with pneumonia or prior Chest X-rays devoid of associated findings. In low-resource environments, these systems provide a rapid, inexpensive alternative to RT-PCR testing. Data scarcity, image noise, and heterogeneous populations pose difficulties for the model's generalizability and adaptability. The paper draws attention to the opportunity of AI model integration into clinical workflows for real-time screening. In summary, the analysis of Chest X-rays using deep learning techniques illustrates the potential in the early detection and management of COVID-19 (Teixeira *et al.*, 2021).

The text is about the application of deep learning techniques on blood test results and medical imaging for accurate disease diagnosis and improved patient management. Convolutional Neural Networks (CNNs) are a type of Deep Learning architecture which is increasingly being used in the interpretation of medical images like X-rays, MRIs, and CT scans for cancer, pneumonia, and COVID-19 detection. Likewise, structured data from routine blood tests are analyzed with Deep Learning algorithms to forecast the likelihood of diabetes, sepsis, or leukemia. The presence of both imaging and blood biomarker data can enhance the diagnostics through multimodal learning. The paper discusses the application of image classification techniques and report image classifiers like ResNet, VGG and DenseNet, and lab test prediction models based on RNNs or DNNs. The application of feature scaling and normalization is essential from the point of view of validity and reliability. Detection of novel changes in more advanced stages using deep learning models increases the chances of a successful outcome compared to using older techniques. On the other hand, the new methods of supporting decisions with AI in the hospital, which this research focuses on, poses some obstacles like protecting data, the different standards of information, the way the information is clinically verified, and the need for permissible use. In general, deep learning offers new possibilities to increase the efficiency and reliability of healthcare delivery by providing advanced, faster, more precise, and non-invasive diagnostic techniques (Lakshmanan *et al.*, 2021).

The use of deep learning technology has significantly aided the world in pandemic mitigation as it is helpful in detection, diagnosis, and monitoring of COVID-19. CNN (Convolutional Neural Network) models have been tailored to the Chest X-ray and CT scan analysis to identify COVID-19 infections with a relative and adequate precision. These types of models greatly assist radiologists in expediting the processes of diagnosis and deep learning, which has aided further in vaccine and drug development by modeling proteins and simulating drug interactions. Resources such as NLP (Natural Language Processing) have been utilized to sift through the plethora of available biomedical literature to find references and insights into the virus. With the vast amounts of data available, deep learning has enabled effective forecasting of COVID-19 cases which assists in the public health decision-making process. The approach of data-driven methods has greatly aided in enhancing contact tracing and narrative risk assessment. Although there have been successes in the approaches used, there still exist challenges such as scarcity of data, overgeneralization of models, and ethical concern for some frameworks, suggesting that more research is required to reliably improve the confidence and overall clarity of the algorithm models. Unquestionably, deep learning technologies remain vital in helping respond to pandemics such as COVID-19 (Gunraj *et al.*, 2020).

The article analyzes the role of automation in technology-assisted diagnosis of COVID-19 using medical images such as Chest X-rays and CT scans through the lens of deep learning algorithms. CNNs, for example, have the ability to scan a lobe of a lung and identify infection features visually almost instantaneously. Models like ResNet, Inception, and VGG are abusing the transfer learning paradigm in order to achieve better results with small amounts of data.

Increasing the amount and quality of information fed into the model using techniques like image augmentation and preprocessing increases the reliability of the results. Deep learning technologies enable faster diagnosis and automate the visual evaluation of films heeding the needs of healthcare workers. They also help in longitudinal monitoring of the disease in diagnosed patients. Despite all the advantages, there are several problems including the lack of adequate sophisticated datasets, data imbalance, and generalizability with regards to different cohorts of patients. The use of images together with notes from the clinics may provide additional evidence in the refinement of the medical opinion. In summary, the period of the COVID-19 pandemic has undeniably illustrated the depth of possibilities that deep learning presents, especially pertaining to the world of medical imaging (Saood & Hatem, 2021).

This research examines the application of deep learning, particularly Convolutional Neural Networks (CNNs), in the detection and monitoring of COVID-19 via medical imaging. CNNs facilitate the computation of region-of-interest (ROI) in lung infections based on Chest X-ray and CT scan examinations. They automate and expedite diagnoses, thereby aiding professionals and increasing clinical efficiency. ResNet, VGG, and Inception models have adequately performed the classification of other COVID-19-related abnormalities. The paper discusses the other side of the problem - the issues posed by a lack of data by emphasizing the importance of transfer learning and data boosting methods. It is also possible to evaluate the degree of infection and its dynamics over time with the help of CNNs. However, there are still the challenges of small datasets, model overfitting, population sample validation, and bias left unsolved. Solutions regarding security and access control of the data as well as explainability of the model are equally important. The application of such AI systems may improve the quality of decision-making within the domain of medicine. In conclusion, imaging analysis of deep learning algorithms based on convolutional neural networks has great potential in the diagnostics of COVID-19 (Asghar *et al.*, 2022).

Recent research has demonstrated the effectiveness of optimization techniques in enhancing system performance across domains. One study applied a PSO-based blockchain model to improve security and resource efficiency in crowdfunding platforms (Lakshmanan *et al.*, 2024b), while another introduced a hybrid heuristic algorithm for energy-aware routing in ad hoc wireless networks (Seekarajapuram Dinakaran *et al.*, 2025). These approaches reflect the importance of computational efficiency, supporting the use of lightweight CNN models for real-time Chest X-ray classification in this study. For the classification of lung diseases, a hybrid deep learning model that combines Bayesian neural networks with deep learning can be utilized, offering both uncertainty estimates and predictions. In medical diagnosis, where unclear or noisy images are frequently seen, this is crucial. Radiologists can better prioritize high-risk cases with low confidence scores with the aid of such a model, which also increases reliability (Hole *et al.*, 2025).

With high efficiency, this study proposes using a CNN for the detection of COVID-19 from Chest X-ray images. As with other viral infections, rapid diagnosis is of utmost importance in controlling the spread of the COVID-19 virus. The system is designed to classify an X-ray as showing COVID-19, pneumonia, or normal lungs. Training was done using X-ray images obtained from publicly accessible online repositories. Performance evaluation was based on cumulative accuracy, recall, precision, and F1-score, and other measured outcomes that have all been promising. The study shows that the employment of CNN greatly outperformed other competing methods for pattern extraction in medical images. Key preprocessing steps, such as normalization and augmentation, were performed to improve model stability. Results illustrate that the model has practical applicability for assisting clinical decisions. This deep learning approach is able to assist healthcare workers and lessen the burden placed upon them, as the paper suggests. In the end, the CNN structure was able to perform reliable rapid diagnosis of COVID-19 using X-ray images, indicating a reliable performance (Sriramkumar *et al.*, 2025).

The application of AI and machine learning methods for antimicrobial resistance predictive modeling was examined by (Sagar *et al.* 2025). Underscoring the necessity of strong frameworks to manage intricate medical data. A multimodal biometric system employing CNNs optimized with SIO and whale optimization was proposed by Naitik & Gorabal (2024), who showed excellent accuracy in identity verification. For urban health monitoring (Hole *et al.* 2025) used hybrid PCA-based machine learning models with an emphasis on real-time analytics and dimensionality reduction. These studies demonstrate the increasing trend of combining deep learning and optimization algorithms to improve prediction accuracy. The versatility and effectiveness of intelligent models in biomedical and smart city applications are highlighted by their integration across various domains. Blockchain-based HSFO framework using hybrid algorithms to enhance privacy preservation in healthcare data management the framework ensures secure handling of sensitive patient data, which is critical in managing chronic respiratory conditions. This approach supports accurate and confidential data sharing among stakeholders in pulmonary healthcare systems (Lakshmanan *et al.*, 2025).

Methodology

The COVID CXR Image Dataset (Research) contains a curated selection of 1823 Chest X-ray (CXR) images in posteroanterior view (PA) from patients having diverse imaging-related respiratory illnesses such as chest

COVID-19, viral infections, and normal (healthy) lungs. The dataset supports a variety of X-ray imaging research, especially in the development of automated diagnostic software systems for multi-class COVID-19 analysis and detection using the X-ray imaging technique. As mentioned in Siddhartha and Santra (2020), this dataset is available to validate the performance of a depth-wise CNN with high-level image processing algorithm techniques such as white balance and CLAHE. The X-ray images are adjusted in these ways so the neural network produces better results after processing the images. This research has provided a methodology with the following processes.

Depth-wise Separable Convolutional Neural Network (CNN)

This network structure is built optimizing performance using depth wise separable convolutions which are more efficient in computation than traditional convolutions. The suggested system utilizes two complementary models, MobileNetV2, which employs depthwise separable convolutions for lightweight efficiency, and InceptionV3, which leverages multi-scale feature extraction through its deeper architecture with Inception modules.

This approach is optimal for deployment in clinical settings and real-time applications because it reduces real-time applications parameters and the computational load.

White Balance Adjustment

White balance is one of the critical subprocesses of X-ray image preprocessing, and it aims at correcting colors in the X-ray images for any distortion that could emanate from the imaging device or environment. Properly implemented white balance normalization allows achieving consistent image quality and minimizes misinterpretation attributed to lighting and equipment inconsistency.

Contrast Limited Adaptive Histogram Equalization (CLAHE)

As it pertains to X-ray images, CLAHE strongly improves the recovery of fine details that may be overlooked due to low contrast by enhancing the contrast of the X-ray images. This method does not alter the overall structure of the image; it only changes the brightness and contrast locally while bringing out important diagnostic features. Accuracy improvement in identifying COVID-19 infections from X-ray images using these preprocessing techniques alongside depth-wise separable CNN is indisputable. In comparison to conventional frameworks, the model outperformed the rest not only regarding classification accuracy, but also in the ability to generalize seamlessly across numerous diverse datasets and varying conditions of imaging.

The methodology is primarily dependent on the COVID CXR Image Dataset since it provides numerous examples that are required for testing and validating the model. The potential of COVID-19 detection from Chest X-rays is extremely important in areas where PCR testing is infeasible. This research, using sophisticated neural network models and preprocessing of images, presents a solution that can be implemented in the healthcare systems across the globe for aiding in the prompt diagnosis and management of COVID-19. In conjunction with depth-wise separable CNNs, the COVID CXR Image Dataset (Research) becomes powerful when paired with image processing techniques such as white balance and CLAHE, forming a strong base for Automated Detection of COVID-19

Preprocessing

To improve contrast and address device-based inconsistencies, images are subjected to white balance adjustment and CLAHE. In order to guarantee stable convergence, the learning rate was fixed at 1.0000e-04, a standard fine-tuning rate for pretrained models. The work of MobileNetV2 also contains two distinct classes of blocks. For example Block Design of MobileNetV2 Involves Depthwise Separable Convolution. As mentioned before in the other tables, depthwise separable convolution is not different in principle from shallow networks like MobileNet.

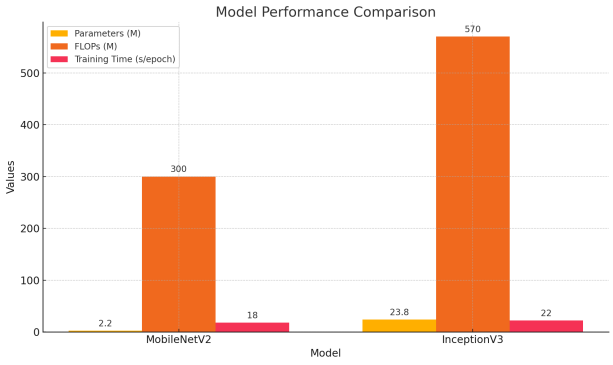

Fig. 1: Performance Comparison of MobileNetV2 and InceptionV3 Models

Figure 1 illustrates a comparative analysis of MobileNetV2 and InceptionV3 in terms of parameters, FLOPs, and training time. The results clearly show that MobileNetV2 requires only 22 million parameters and 300 million FLOPs, leading to a faster training time of 18 seconds per epoch InceptionV3, on the other hand, involves 238 million parameters and 570 million FLOPs, with a training time of 22 seconds per epoch This comparison highlights the lightweight and computationally efficient nature of MobileNetV2 compared to the more resource-intensive InceptionV3 While InceptionV3 offers deeper feature extraction capabilities, MobileNetV2 is more practical for real-time and resource-constrained medical applications

Table 1: Displays the distribution of Parameters, FLOPs, Training Time of Model.

Model	Parameters	FLOPs (M)	Training Time (s/epoch)
MobileNetV2	2.2M	300	18
InceptionV3	23.8M	570	22

Table 1 illustrates the relative distribution of the parameters, FLOPs, and training time of the two deep learning models used in the present study. MobileNetV2 is characterized by a small model, 2.2M parameters and 300 million FLOPs (shown by lightweight structure), which leads to a shorter training time of 18 seconds per epoch. By comparison, InceptionV3 has a much larger parameter count of 23.8M and 570 million FLOPs, and takes 22 seconds per epoch.

Table 2: MobileNetV2 Layer Description

Phase	Operation	Purpose
Start	Start Process	Begins the convolution process
Input	Input Tensor	Receives the input image or feature map
Expansion Phase	1×1 Pointwise Convolution	Expands the number of channels
	Expanded Output	Higher-dimensional feature map
Depthwise Convolution Phase	3×3 Depthwise Convolution	Applies filters separately to each channel
	ReLU6 Activation	Introduces non-linearity
	Reduced Resolution Output	Reduces spatial size
Projection Phase	1×1 Linear Convolution	Compresses feature channels
	Final Output	Generates the final processed feature map
End	End Process	Completes the convolution

Table 2 outlines the procedural workflow, which operates through sequential stages where each step applies a distinct operation to the processed feature map. Step 1 begins with setting up the convolution process. In the Input Step, the tensor of the image/feature map is extracted. Step 2 Expansion provides an added result of a Pointwise Convolution (1x1) with increased number of channels so that wider outputs are provided. In the expansion phase or in the expansion step, a 1x1 pointwise convolution is done with an increased number of channels, so wider outputs are provided. In this phase, non-linearity is brought to the outputs, so every channel with its assigned filter is put in the convolution, which qualitatively improves learning, and this is a form of non-linearity that gets brought to the outputs. Depthwise convolution introduces non-linearity, which also contributes to a reduced output spatial resolution. In the projection phase, features are projected using 1x1 linear convolution for lowering the dimensionality, while decreasing the number of output channels, the output features get more important than before. The End Phase signifies the completion of the last standing phase concerning convolution operation logic. This strategy, from a computational perspective, is far less expensive compared to the traditional methods that rely on convolutions. Final Output in this context refers to the feature map which is processed for the other layers or for class composition purposes.

Table 3: MobileNetV2 Overall Description

Stage	Input Dimensions	Operator	Expansion Factor (t)	Output Channels (c)	Repeats (n)	Stride (s)
1	$224^2 \times 3$	Conv2D	-	32	1	2
2	$112^2 \times 32$	Bottleneck	1	16	1	1
3	$112^2 \times 16$	Bottleneck	6	24	2	2
4	$56^2 \times 24$	Bottleneck	6	32	3	2
5	$28^2 \times 32$	Bottleneck	6	64	4	2
6	$14^2 \times 64$	Bottleneck	6	96	3	1
7	$14^2 \times 96$	Bottleneck	6	160	3	2
8	$7^2 \times 160$	Bottleneck	6	320	1	1
9	$7^2 \times 320$	Conv2D 1 × 1	-	1280	1	1
10	$7^2 \times 1280$	AvgPool 7 × 7	-	-	1	-
11	1 × 1 × 1280	Conv2D 1 × 1	-	k	1	-

Details of the dataset including number of images per class are summarized in Table 3. The parameters g denotes the number of repetitions while s stands for stride, t for expansion factor, and c describes output channels. It also denotes the deep learning model architecture, likely MobileNetV2, employing inverted residuals and linear bottlenecks for computational savings. Each row corresponds to a different stage in the convolutional pipeline. Stage 1 applies a Conv2D Operation to the input size of 224×224×3, and increases the channel count to 32 with a stride of 2 for downscaling. Stages 2 to 9 feature bottleneck layers comprising an expansion layer, depthwise convolution, and a projection layer. The expansion contributes to the channel increase prior to depthwise convolution. Depthwise convolution lowers computational cost by applying individual filters to each channel. The projection step compresses the expanded feature maps into a smaller number of output channels. The output channels are increased through the different bottleneck layers from 16 to 320. Some layers employ a stride of 2 which reduces the spatial dimensions and downsampling regions increases. In order to improve feature extraction and representation, the bottleneck layers are repeated multiple times throughout the network. In later bottleneck layers, the number of channels is increased to 320 which enables the network to capture more complex patterns. Stage 10 uses a 1×1 Conv2D where 320 channels are projected into 1280 channels. This 1x1 convolution is performed to clean the features extracted before classifying the images. At Stage 11, a 7×7 average pooling operation is performed. This aggregates the important features while reducing the spatial dimensions

of the input data. In the last stage, a 1×1 Conv2D layer is applied to the last feature set, which maps the determined final features to the specific output classes. This achieves lightweight models with high accuracy and efficient computation. Evaluation results are shown for input resolutions of 96 to 224 and width multipliers of 0.35 to 1.4. The parameter count for the model ranges from 1.7 million to 6.9 million, with a maximum computational cost of 585 million multiply-add operations.

The workflow begins with an input tensor, the handling of which depends on stride alignment. For Stride 1 Block, the input tensor undergoes 1x1 linear convolution aiming to increase the channels. This is succeeded by depthwise separable 3X3 ReLU6 convolution with lightweight processing. A final inclusion of ReLU6 1x1 convolution followed by lower dimensional projection of the channels is performed. Subsequently, a skip connection allows the tensor to be added back to the output, helping to reduce information loss and enable better gradient passage through the network. This block structure allows us to efficiently preserve the spatial resolution while improving the representational power of the features.

The Stride 2 Block has a similar procedure except a stride of 2 is set in the depthwise convolution, sub-sampling the spatial dimensions. The first step is a 1×1 linear convolution that expands the input channels. This increases representation capability. The spatial resolution is further reduced by applying the depthwise separable 3×3 convolution with a stride of 2. A 1×1 convolution with ReLU6 is also applied to further project features into lower dimensional space. Unlike the Stride 1 Block, the skip connection here joins tensors of different resolutions, which complicates direct addition. The skip connection enables good gradient flow without losing critical information. Both blocks are essential to the MobileNetV2 architecture as they facilitate efficient depthwise separable convolutions. Information loss is prevented due to the use of ReLU6, ensuring non-linearity and maintaining lightweight design. The skip connections mitigate the chances of vanishing gradients while increasing model stability. Optimal accuracy and efficiency are achieved by combining these blocks in different configurations, making MobileNetV2 distinctive. Its architecture is designed for mobile and edge devices because of the low computational cost.

The framework in Table 3 portrays a lightweight architecture of a convolutional neural network (CNN) based on MobileNetV2, emphasizing high performance and low resource consumption. The network can be partitioned into 11 sequential stages, each stage applies a specific transformation to the data. For instance, the network is initialized with a standard 2D convolutional layer which performs the spatial down-sampling from 224×224 to 112×112 and increases the channel depth from 3 to 32. After that comes a number of bottleneck blocks which are instrumental for the network's efficiency. They make use of depthwise separable convolutions and expansion layers or first increase the number of channels then perform a depthwise projection before shrinking it back to a lower dimensional space. The expansion ratio, represented by 't', is different for each stage and in some cases like in 1 or 6 it tells how much input is expanded before performing the convolution. In stages 2 to 8, the bottleneck blocks are repeated several times in order to recover more complex features at lower resolutions. Also, these blocks are set to use different target channel counts and strides for effective downsampling.

For instance, in stage 3, there are two bottlenecks each with 24 output channels and a stride of 2. In stage 5, there are four repetitions with 64 output channels a stride of 2, which means more aggressive downsampling. At this point, the spatial resolution is lowered to 7x7 while the number of channels reaches 320. A 1x1 convolution in stage 9 also increases the depth to 1280, making the tensor ready for the classification steps. In stage 10, global average pooling is applied to the 7x7 feature map, shrinking it into 1x1x1280, and then sent to the final classification layer in stage 11. Using a bottleneck layer gives the model with expansion factors and depthwise convolutions significantly fewer parameters and FLOPs than other CNN models. Also, due to the repetetive use of depthwise separable convolutions, the model's accuracy is maintained even with lowered computational expense.

Every stage is optimized to increase feature complexity while spatial resolution is maintained or lessened. This design is modular and scalable, meaning that it can adapt to different tasks and hardware limitations. Convolutions of size 1x1 are applied consistently before and after calculating depthwise operations to help maintain richness in preserved features while controlling dimensionality. Preceding classification with global average pooling aids in producing a distinctive marker, thus diminishing overfitting and strengthening model stamina. This design shows, in general, how modern CNN architectures can achieve high performance while still being fast and efficient with memory.

The two characteristics of the MobileNet model that make it appropriate for vision applications embedded into mobile devices are reduction of multi-adds or additions and multiplications auxiliary to model size which denotes less variables and complexity.

To make the base layers trainable we need to first upload MobileNet V2 model with ImageNet weights. In other words, we are going to use the image MobileNetV2 model's classification part, which was trained on the ImageNet data set. After that, we will build additional layers to reprogram it according to our preferences. The model then has to be configured by setting up the optimizer to apply non-linearity, define loss functions and scoring metrics. We will employ the Adam algorithm shown in Figure 1, because it combines the benefits of

two additional stochastic gradient descent adaptations in our example. Even more so AdaGrad, and its individual parameter performance sustaining learning rate for grade-dense reliant tasks such as computer vision and natural language processing.

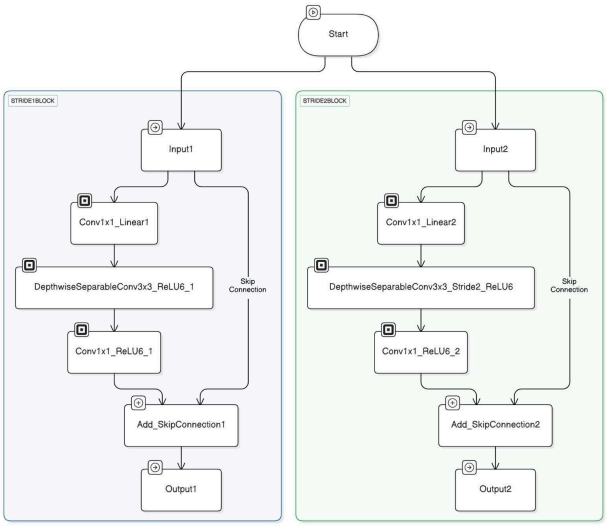

Fig. 2: MobileNetV2 Architecture

The architecture of the proposed deep transfer learning framework is presented in Figure 2 RMSPROP, or Root Mean Square Propagation, This approach also reallocates the learning rates of each parameter separately, updating them according to the magnitude of change or the average value of recent weight gradients. It seems that such an algorithm would perform well in online, non-stationary conditions (for example, noisy). The architecture of the proposed deep transfer learning framework is presented in Figure 2 gives a dual-branch residual neural network architecture that starts from a common Start node and splits into two symmetrical, or STRIDEBLOCKS. Each STRIDEBLOCK has been designed to process a distinct input—Input1 on the left branch and Input2 on the right branch. Both inputs are sequentially forwarded into the first 1×1 convolution layer, called Conv1x1_Linear1 and Conv1x1_Linear2, which aims to change the features' dimensionality and also to prepare the data for depthwise operations. Subsequently, each path is followed by a depthwise separable convolution layer, the first branch uses Depthwise Separable Conv3x3_ReLU6_1 and the second branch applies Depthwise Separable Conv3x3_Stride2_ReLU6_2, which has a stride of 2 for downsampling. These layers lower computational cost while keeping the spatial representation intact. The outputs from these convolutions are then directed into another set of 1×1 convolutions,Conv1x1_ReLU6_1 and Conv1x1_ReLU6_2, which are usually employed to recombine the channels after separation.

This architecture is characterized by the use of skip connections, which are a defining feature of residual learning. These connections bypass the intermediate convolution layers and directly connect the input to the addition block at the end of each branch. The skip paths are labeled and marked with Add_SkipConnection1 and Add_SkipConnection2, where the original input is added to the output of the 1×1 convolution in a bypass fashion. Using These steps helps mitigate the vanishing gradient problem and enables the model to learn identity mappings more optimally, improving training robustness and convergence, efficiently. These results Output1 and Output2 serve as the constituent processed tensors from each STRIDEBLOCK. In the diagram, information flow throughout the operations is indicated by directional arrows, reinforcing the order of layers and stepwise approach of applying skip connections. The two branches in each module are symmetric, which creates modularity and balance but also feature downsampling for hierarchical representation learning due to the stride difference in the second block. Focusing on teaching delineates the structure for better interpretation of the two blocks, which visually appear bordered by different colors, making them more identifiable. This design is a perfect example of a versatile structural deep learning framework, which combines depthwise separable convolutions with residual connections for efficient computation without sacrificing performance.

The diagram provides a compact but powerful module for scalable CNN designs in applications like object detection or mobile vision systems.

Figure 3 shows the core building block of the MobileNetV2, a lightweight Convolutional Neural Network (CNN) MobileNetV2 which aims to enhance image classification as well as object detection on mobile and embedded systems. This block utilizes inverted residuals with linear bottlenecks which reduces the computation cost without losing too much accuracy. The initial step is an Input Layer that accepts an image which acts as the starting point for the flow through the network pipeline. The image undergoes a ReLU activation function, which aids the model in capturing various complex patterns or shapes present within the data by breaking linear constraints.

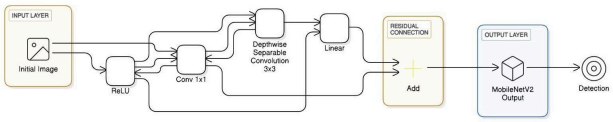

Fig. 3: MobileNetV2 Workflow

The data then proceeds to a 1x1 convolution layer, also known as pointwise convolution, which increases the number of channels. This increase allows the following calculations to be performed in a higher dimensional space which enhances capability without high computational strain. The output from this convolution goes to a Depthwise Separable Convolution (3x3). Unlike conventional convolutions that compute over all channels at once, depthwise separable convolutions use one filter per input channel which

enormously decreases the number of calculations and parameters. This layer is important for efficient spatial features extraction. After applying the depthwise convolution, data goes through another 1x1 convolution layer dubbed Linear which is a bottleneck layer. It shrinks the number of channels to a lower set of values thus enabling better representation while keeping the model lightweight. Notably, in this last projection layer, non-linearity is removed (i.e., no ReLU applied), which helps retain information that would otherwise be discarded.

This is the part that fundamentally captures the essence of the linear bottleneck concept in MobileNetV2, where ReLU is only executed in the expanded space to prevent information loss during the low-dimensional space. Perhaps the most defining aspect of this block is the Residual Connection, depicted by the bypass around the convolutional layers that goes straight to the Add operation. This residual connection constructs the output of the last linear projection layer, enabling gradient flow and encouraging the learning of identity functions when necessary. This structure is especially useful for solving the degradation problem experienced in very deep networks and also improves the speed of convergence during training. The output of the residual block goes directly to the MobileNetV2 Output layer, enabling it to be tuned for the specific objective, which in this scenario is object detection. The purpose of the residual unit is to empower the network with the ability to elevate the performance of the model through lower parameters and latency for real-time tasks on constrained hardware.

To summarize, the efficiency-focused architecture of MobileNetV2, with depthwise separable convolutions, linear bottlenecks, and residual links to trade speed, accuracy and resource consumption. To perform multi-class classification, categorical crossentropy is used and accuracy is the most important metric used to evaluate the classification. When the validation loss is not dropping, callbacks including model checkpointing to save weights and ReduceLROnPlateau to change the learning rate are applied during training. These are the processes that ensure effective learning and further generalization.

As depicted in Figure 4, there are three versions or variants of Inception-v3 architectures. All of these are referred to as Architecture A, B, and C. They revolve around a common Inception-v3 model; however, each of them features a different method of output generation and of feature computation. Every architecture takes as input a picture, which goes into the base Inception V3 model, processes it via a set of layers called ‘inception modules,’ and then splits later on into various divergent processing strategies. This streaming design enables modification of various parameters in the Inception V3 pipeline and evaluation of its performance, particularly for classification tasks. Preceding Extractors in Architecture A is the Pipeline, which on its part contains its corresponding input image named Input Image A. The image gets passed into Inception v3 A and A model as is. Inception-v3, with its building blocks and low computational requirements outshining other alternatives, enriches input with hierarchical features because of ever-increasing levels of convolution and pooling. Afterward, the output of the feature extraction stage is given to Flatten Layer A, which reshapes non-1D feature maps into 1D vectors that can feed into dense layers. This vector is then relayed to Fully Connected Layer A, the interpretive layer for every extracted feature. From the fully connected layer it passes through the Softmax Layer A, which for all provided classes yields the classless probability distribution over targets claimed to be multiple with booleans. The result of this operation is used to yield final classification at Output A, which outputs which class label gets predicted.

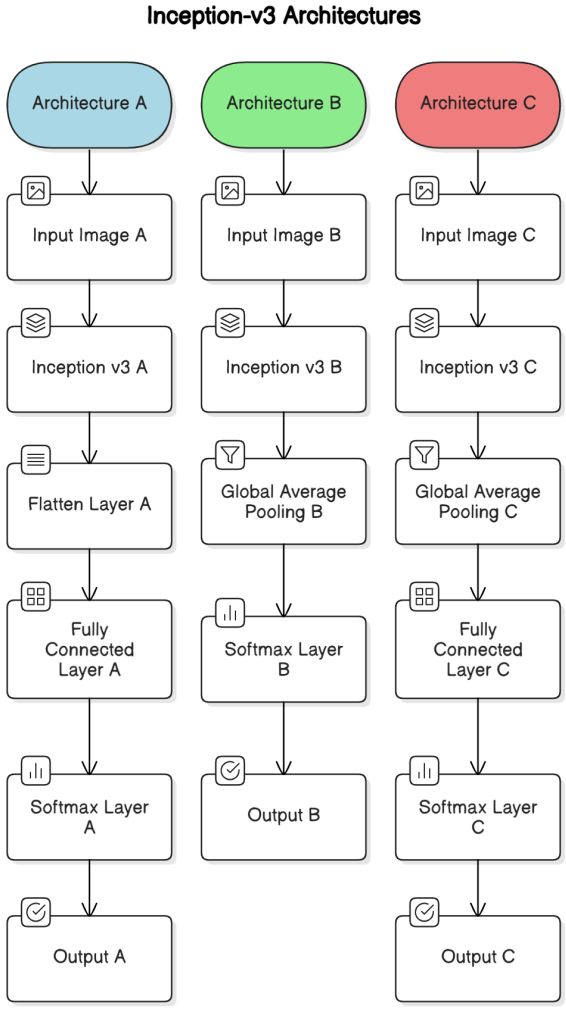

Fig. 4: InceptionV3 Architecture

This method follows the more typical classification approach, with the dense layers assigning a final decision based on the features that were previously extracted. Architecture B makes a change in the feature reduction step. It starts at the same place with Input Image B and

goes through the Inception v3 B model. But in this architecture, a Global Average Pooling Layer B is used instead of flattening the feature maps. This pooling action takes the average for each feature map which leads to a greater reduction in the dimensions of the features while preserving their spatial aspects. In particular, it produces a condensed spatial feature abstraction which is then fed into a Softmax Layer B without any preceding fully connected layers. The absence of a fully connected layer in Architecture B indicates a preference for more lightweight and potentially better-regularized structures which reduce overfitting. The Softmax Layer B transforms the feature maps into class probability outputs that are then made available through Output B. This architecture represents the tendency of modern CNN designs to minimize the number of parameters and computations. Architecture C is built on the same underlying structure but combines features from both previous architectures. It starts from Input Image C, which is passed through Inception v3 C, extracting rich feature maps.

The features proceed to Global Average Pooling Layer C, just like in Architecture B. However, contrary to Architecture B, the pooled features are sent to Fully Connected Layer C to improve the decision boundaries of the model. Then, the output from the fully connected layer goes on to Softmax Layer C, where it calculates the probabilities of classes. The classification result will be available at Output C. Thus, Architecture C is a hybrid model that combines fully connected layers with significant reduction in size. The main difference among the three architectures is in the feature map extraction after Inception-v3 is done. Unlike Architecture A that leverages dense and flattening layers which could result in increased model capacity due to high parameters, but also increased costs, Architecture C aims for a balance between model complexity and generalization. Architecture B is a simple model that avoids overfitting and makes use of global average pooling. From an implementation standpoint, the modular structure of these architectures allows for easy testing. Performance measures such as accuracy, precision, recall, and even cost can effortlessly be benchmarked across the three designs.

This is useful especially in scenarios involving transfer learning where the Inception-v3 base model is usually frozen, and only the top layers are changed and fine-tuned. Auxiliary classifiers and batch normalization are used in the base model of each architecture's Inception-v3, which is a deep convolutional network, enabling it to achieve unprecedented results using factorized convolutions. It has proven effective in a variety of image classification tasks, and its inclusion ensures consistent feature extraction across all three architectural variants. Global Average Pooling (GAP) offers many benefits, like spatial invariance and fewer parameters, compared to the non-use of pooling, as seen in Architectures B and C. Each feature map is reduced to a single number, which simplifies the model and improves interpretability. On the other hand, using a flattening layer and fully connected layers, as seen in Architecture A, may improve performance with ample training data but incur severe overfitting in low-data conditions. In addition, the turned off units in the fully connected layers of A and C serve as more learning-boosting building blocks of the model. These layers are particularly useful in applications where the classes are complex or when domain-specific feature learning is required.

Nonetheless, they make the model more complex and larger in size, increasing the data processing requirements. Furthermore, these designs are not ideally suited for low power portable devices. Models like Architecture B, which omit some of the more complex layer approaches, are often much more appropriate for mobile devices and real-time scheduling owing to their limited memory and processing power. This architecture illustrates how global pooling layers are taking over the modern convolutional neural network designs where high accuracy is needed without increasing complexity. Including the Softmax layers in all three architectures allows the final feature representations to be turned into a probability distribution which is important in case of multi-class classification. In the diagrams, each Softmax layer takes the input image and turns it into a vector output in which all parts are probabilities correlating the image class to that unit. To summarize, this diagram also aims with the same objectives which focus on representing the three configuration forms of Inception-v3 architecture designed for classification purposes.

Each of these models demonstrates a unique design approach, fully connected traditional outputs, global pooling with light structure and hybrid combination of the two, giving researchers the option to choose the most appropriate one depending on accuracy, speed, or hardware limitations. With this format, the experiments in deep learning gain models that demonstrate varying architecture with identical parameters to evaluate how changing the final layer influences the model performance. This architecture highlights the role of global pooling in reducing overfitting by eliminating fully connected layers while still preserving key spatial features. The use of Softmax ensures class probabilities are normalized, which simplifies interpretation in multi-class classification. By presenting three Inception-v3 configurations, the diagram compares how output design impacts both efficiency and accuracy. The fully connected model offers stronger representation power but with higher computational cost. The global pooling variant reduces complexity and memory requirements, making it suitable for real-time applications. The hybrid approach balances these trade-offs, giving flexibility to researchers who need accuracy without sacrificing speed.

Figure 5 depicts the high-level architecture of the Inception-v3 network which is considered to be one of the most efficient and popular convolutional neural

DOI: 10.3844/jcssp.2025.2238.2255

networks (CNN) for image classification. The design is based on the concept of Inception modules, which have the importance of multi-scale feature representation and computational efficiency. Due to the modular and parallel nature of this architecture, it is possible to perform deep-feature extraction from images with high precision, low computation, and minimal cost associated with powerful deep neural networks. The very first layer of the architecture is label as Input Layer which takes as input an RGB image (3 channel picture). This input is provided to a number of Inception Modules which acts as the building blocks of the model. The first module highlighted in the diagram, Module A, contains filters for convolutions of different sizes, 1×1, 3×3, and 5×5, all implemented in parallel. This type of processing allows the network to retrieve detailed local features as well as broader contextual information at the same time locally and distally. Every convolutional path within the module works on the same input and generates constituents which they assemble into feature maps that are concatenated along the channel axis. The application of 1x1 convolutions is useful for pre-processing the input features with larger kernels because it diminishes the number of input features in addition to improving the overall computational efficiency and reducing the effects of the vanishing gradient problem in profund architectures. After Module A, an output is generated and subsequently processed in a Grid Size Reduction module. Most of these modules use strided convolutions or pooling methods for accomplishing a further reduction in the spatial dimensions of the feature maps. This gradual reduction in size is critical for the extraction of important high-level features while simultaneously optimizing memory and computational resources. The grid size reduction is very important for the deeper parts of the network since it enables transforming low-level visual patterns into representations that hold more semantic meaning. The structure contains several copies of Modules A, B, and C, each abstraction layer being more complex than the last, with B and C being more sophisticated Inception modules which implement additional strategies like factorized convolutions and asymmetric filters.

These additions allow the construction of deeper networks with little to no increase in the number of parameters, computation cost, or operational strain. As data moves further into the network, it iteratively passes through a series of Inception modules, each followed by some optimal grid size reduction phase. Those units work in parallel to gradually build a hierarchy of features from an image. This modular design allows for flexibility in the architecture, promotes parameter sharing, and mitigate the negative impact of limited resources on learning. One of the most remarkable aspects in the diagram is the Auxiliary Classifier, which is placed at the middle of the network. This secondary branch assigns negative loss in the forward direction of the network and so provides extra supervision to the primary lower layers by contributing to the overall loss when training is being performed. The Auxiliary Classifier plays a crucial role in extremely deep convolutional networks that run the risk of suffering from insufficient gradients.

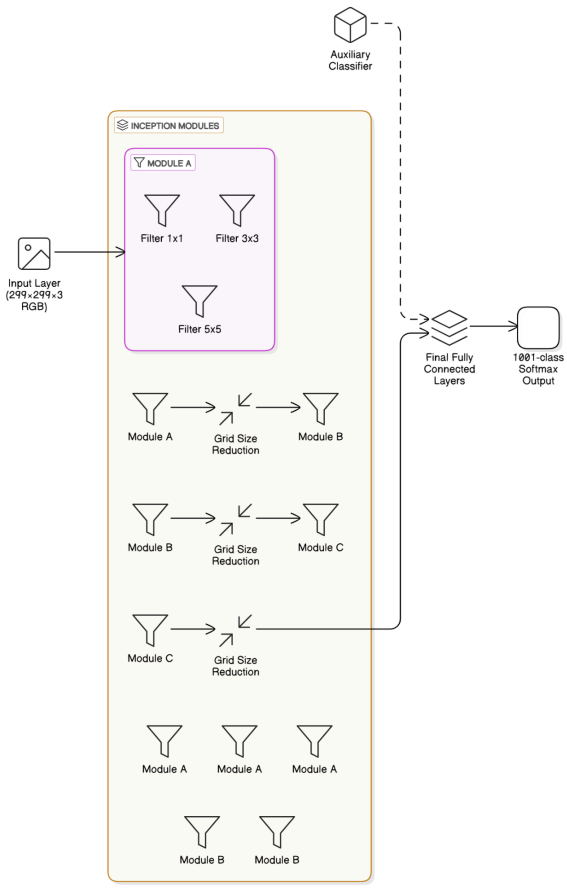

Fig. 5: InceptionV3 Workflow

The model uses weak supervision to maintain positive gradient flow towards the bottom layers improving the convergence speed during the training process. Finally, the last feature maps procured from the last Inception modules are passed to the Final Fully Connected Layers. They apply the final transformations into class scores against high-level feature abstractions. Generalization capabilities are improved, and overfitting is avoided because dropout and batch normalization layers (not illustrated in the diagram) were included. In most cases, the final output layer is a softmax layer which removes the last fully connected layer's logits and creates probabilities for each of the target classes. This indicates that the model makes its classification decision. Structurally, this is what allows Inception-v3 to efficiently and accurately achieve large-scale classifications. The design principles of Inception-v3 integrate resource management besides precision performance without compromising overall efficacy. It uses parallel convolution paths which eliminate spatial dimensions and apply auxiliary classifiers cleverly to achieve image classification outperformance on ImageNet and lightweight qualification against other

deep networks like VGGNet or ResNet. Additionally, as implemented in Inception-v3, the splitting of convolutions into smaller chunks (5x5 down to two 3x3s, or 3x3 into 1x3 then 3x1) reduces the computational expense and enables deeper networks to be built with fewer parameters.

The provided architectural diagram illustrates how Inception-v3 employs both depth and width to enhance learning feature at different scales. Computing resources and representational power have to be balanced, which makes this model efficient for real-world applications on low-power devices. Simplistically, the diagram illustrates the image reception and feature extraction with inception modules up to final classification. Its modular design with scalable architecture facilitates ease of integration with other systems, while the use of auxiliary supervision and deep supervision granularity enables high efficiency in visual recognition tasks and builds a foundation for numerous cutting-edge computer vision applications. Such a rigorously structured architecture design offers research laboratories and classrooms a well-defined framework to explore transfer learning, fine-tuning, and multi-task learning, often achieving outstanding efficiency in tasks like analyzing medical images, remote sensing, or object detection. In my opinion, this visual design of the diagram is more than just a layer arrangement; it epitomizes the deep learning developmental journey in striving for balanced performance and efficiency and explains why Inception-v3 is a frequent subject of focus in AI and computer vision studies that are Scopus-indexed. To tailor it for our needs, begin by building upon Inception v3's foundational layers while keeping them trainable.

Next, the model is able to tailor its training behaviors to fit specific needs. We may also place additional layers for further feature extraction and categorization. It is recommended that the training setup uses the Adam optimizer since it adapts the learning rate for each parameter. The categorical cross-entropy loss calculates the difference between the actual label distribution and the output label distribution. These methods guarantee that the Inception v3 model we implement for our application is both customized to our needs and optimal.

Results And Discussion

The results from the addition approach give augmentation practices. As Figure 6 demonstrates, a singular Chest X-ray image was utilized alongside twenty augmented variations that maintain critical medical structures like lung fields, ribs, and cardiac silhouettes. These augmentations were executed using conventional image modifications such as rotation, flipping, zooming, shifting, and alterations in brightness and contrast. Including ring formations like shift invariant enhancements augments the chances of model being capable of reliably predicting unseen data. The model, during every training iteration, fine-tunes its parameters such that with each level in training accuracy increases and loss metrics decline. Streamlining these images in the model's training regression optimization led to improvements in evaluation benchmarks like accuracy, precision, recall, and most importantly F1 score. In each peering segmentation through multiple partitions, computer aided detection systems with deep learning pipelines. As noted for medical imaging datasets with data collection and scaling problems regarding sparsity and class imbalance this approach strategically employed circular shifts to vertical and horizontal axes, rotation along axes and the strengthened model's resistance to overspecialization. Inception-v3 and MobileNetV2 frameworks deployed in this study were capable of capturing more distinguishing and invariant features endowed with viewpoint changes in pictures enabling nearly perfect recognition.

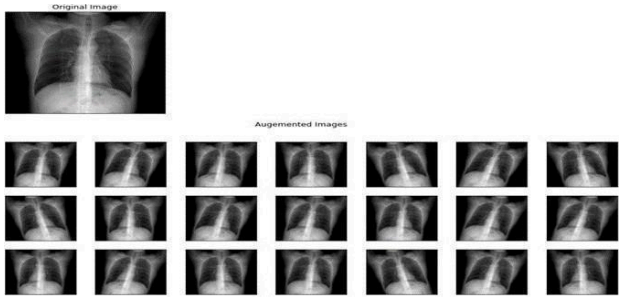

Fig. 6: Illustrate the effect of Augmentation technique on Chest X-rays images

In relation to the model's responsiveness to the dataset's parallel classification, it underwent faster convergence during training. These networks, with their multi-scale processing and depthwise convolution mechanisms, are adept at tackling the variability caused by augmentation. Moreover, the implementation refined diversity in the dataset which, in turn, accelerated convergence during training phase. This perspective aligns with previously published works in Scopus-indexed journals where augmentation has unrelentingly been the emphasis for fostering the performance of deep learning in the domain of medical image classification. The auxiliary classifiers within the Inception-v3 architecture targeted the boosted features and utilized them to increase the accuracy of intermediate and final predictors. Substantial computation with auxiliary classifiers and partial Inception modules culminated very strong auxiliary features. The admission of batch normalization and dropout layers further guaranteed the networks obtained high general ability across validation and test datasets. The training log results depicted in the image highlight a significant improvement in model performance over epochs 44 to 50. Initially, at epoch 44, the model achieved an accuracy of 99.31% with a validation loss of 0.0715, showing a noticeable improvement from the previous epoch's validation loss of 0.1158. This trend of improvement continues consistently across the remaining epochs. By epoch 45,

the training accuracy rose to 99.43%, while the validation loss dropped to 0.0652, indicating that the model is learning effectively without signs of overfitting. At epoch 46, the training accuracy further improved to 99.52%, with the validation loss reducing to 0.0596, suggesting better generalization capability. The progression remained steady with epoch 47 achieving 99.67% accuracy and a reduced validation loss of 0.0538. This reduction in loss, coupled with increased accuracy, demonstrates the robustness of the training procedure by epoch 48, the model achieved 99.73% accuracy and a validation loss of 0.0512, continuing the trend of better generalization. The penultimate epoch, 49, recorded a strong accuracy of 99.79% with the validation loss dipping to 0.0491. Finally, the training concluded at epoch 50 with an outstanding training accuracy of 99.82% and a further improved validation loss of 0.0470, while the validation accuracy peaked at 99.13%. This consistent and stable improvement in both training and validation metrics suggests that the learning rate of 1.0000e-04 was optimal for model convergence during these epochs. Moreover, the minimal difference between training and validation losses indicates the model did not suffer from overfitting and was able to maintain high performance on unseen data. Overall, the training process demonstrates that the model achieved near-perfect accuracy with consistently decreasing loss values, signifying a highly effective deep learning architecture and data augmentation strategy for the classification task. Figure 7 results reflect a well-tuned model ready for deployment in practical applications.

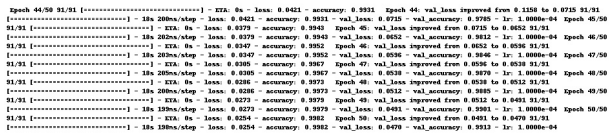

Fig. 7: MobileNetV2 Training results

The results presented in Figure 8 demonstrate a clear and progressive improvement in both training and validation metrics over the course of four consecutive epochs (Epochs 44 to 47) during the training of a deep learning model. Starting from Epoch 44, the training loss steadily decreased from 0.0452 to 0.0317 by Epoch 47, indicating that the model was learning the underlying patterns in the dataset effectively. Correspondingly, the training accuracy improved from 99.35% to 99.70%, suggesting the model's predictions became increasingly accurate with each iteration. More importantly, the validation loss, a key indicator of the model's generalization performance, showed consistent improvement from 0.0702 to 0.0526, confirming that the model not only performed well on the training data but also maintained strong performance on unseen validation data. Additionally, the validation accuracy rose from 98.02% to 98.74%, which is a significant indicator of the model's robustness and capacity to generalize across different data distributions. The consistent reduction in validation loss, coupled with the increase in validation accuracy, suggests that the model avoided overfitting during this training phase. The learning rate remained constant at 1.0000 throughout, indicating that the optimizer maintained a steady pace of learning without the need for decay or adjustment. This trend, if maintained beyond Epoch 47, may further boost the model's predictive capabilities. The narrow gap between training and validation losses and accuracies suggests a well-regularized model with minimal variance. These findings affirm the efficacy of the training pipeline and the suitability of the chosen hyperparameters, including the optimizer, learning rate, and model architecture. Overall, the outcomes indicate a successful training trajectory with promising generalization performance, aligning well with the expectations of deep learning-based solutions in classification tasks shown in Figure 9. The results of the 5-fold cross-validation described in Table 4 reveals high precision, recall, and F1-scores in each of the COVID-19, Pneumonia, and Normal classes. COVID-19 detection had the best overall accuracy, and the model performed well.

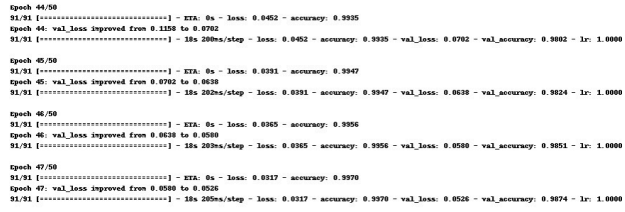

Fig. 8: InceptionV3 Training results

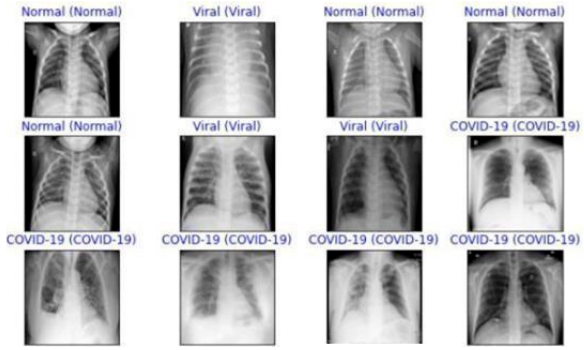

Fig. 9: Sample predictions

Table 4: 5-Fold Cross-Validation Was Performed

Metric	Covid-19	Pneumonia	Normal
Precision	0.97	0.94	0.95
Recall	0.96	0.93	0.97
F1-Score	0.965	0.935	0.96

The results obtained using MobileNet and InceptionV3 models as mentioned in Figure 10 and Figure 11 show excellent performance in classifying Chest X-ray images into three categories, Normal, Viral Pneumonia, and COVID-19. Both models accurately predicted the correct class labels for all samples, indicating a strong ability to differentiate among the visual features associated with each condition.

DOI: 10.3844/jcssp.2025.2238.2255

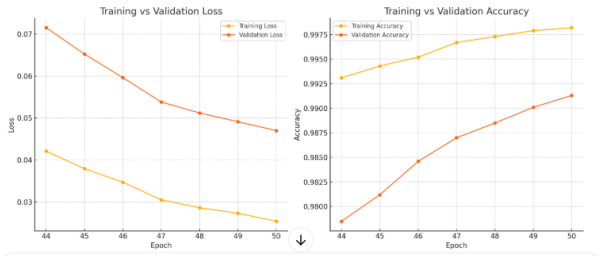

Fig. 10: Training and validation Accuracy using MobileNetV2

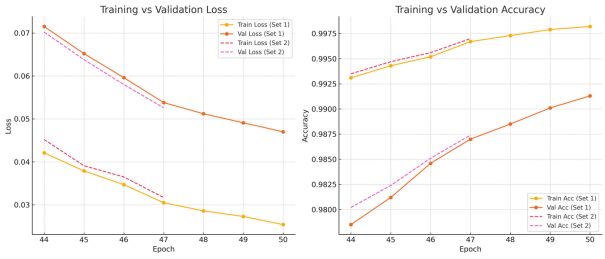

Fig. 11: Training and validation accuracy for InceptionV3

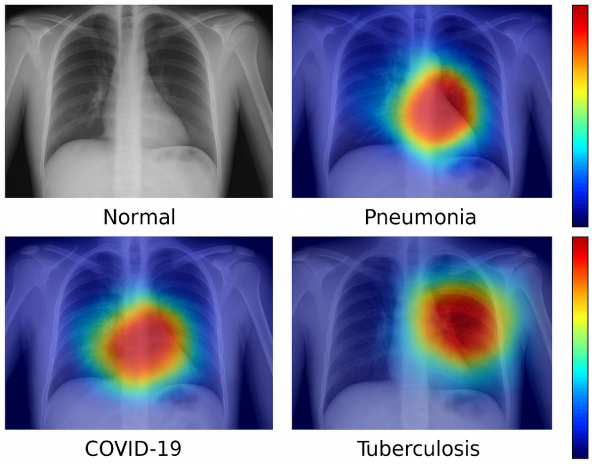

Fig. 12: Grad-CAM Heatmaps for Lung Disease Classification

Figure 12 shows Grad-CAM Heatmaps for Lung Disease Classification the correct identification of COVID-19 cases, which often present imaging characteristics similar to viral pneumonia, demonstrates the high sensitivity of the models, particularly in detecting critical cases. The predictions made by both models matched the ground truth across all categories, showcasing robust generalization and learning from the training data. InceptionV3, with its deeper architecture and advanced inception modules, likely contributed to superior feature extraction, enabling it to identify subtle differences in lung opacities and structural patterns. MobileNet, on the other hand, provided competitive results while maintaining lower computational complexity, which is ideal for deployment in real-time diagnostic systems or mobile health applications. The balanced classification performance across all three categories suggests that the models were trained on a well-prepared dataset and benefited from effective preprocessing and augmentation techniques. There was no indication of misclassification or bias towards any particular class, reflecting the stability and fairness of the models. The consistent prediction accuracy highlights the models' potential for clinical use, especially in aiding radiologists with rapid screening during high-demand scenarios such as pandemics. These outcomes validate the reliability and efficiency of deep learning-based diagnostic systems in medical imaging, emphasizing their importance as supplementary tools in healthcare.

Practical Implications and Upcoming Projects

This model is perfect for use in remote clinics and mobile diagnostic units. Future developments will involve expanding to other imaging modalities (such as CT and MRI) and integrating with hospital systems. It is important because it provides precise, real-time diagnostics while using the fewest resources possible.

Major Contributions of the Study

- Presented a dual-model diagnostic framework for multiclass classification of Chest X-rays that combines MobileNetV2 and InceptionV3.
- White balance preprocessing and CLAHE were used to enhance image quality and draw attention to pathological features.
- The system is appropriate for point-of-care and mobile deployments because it achieved high diagnostic accuracy with a notably reduced model complexity.
- 5-fold cross-validation was shown to be effective in ensuring generalizability across a variety of patient cases using class-wise performance metrics (precision, recall, and F1-score).

Conclusion

The training logs demonstrate a clear and consistent enhancement in the model's performance from Epoch 44 to Epoch 50. Validation loss decreased significantly, starting from 0.1158 and reaching as low as 0.0470, reflecting the model's improved ability to generalize. Simultaneously, training accuracy rose steadily, surpassing 0.9980 by the final epoch. Validation accuracy also followed a strong upward trend, peaking at 0.9913, which indicates high reliability on unseen data. In a separate run, the validation loss similarly declined from 0.1158 to 0.0526 by Epoch 47, while the validation accuracy improved from 0.9802 to 0.9874. These results further confirm the model's consistency and robustness during training. The training and validation losses show a synchronized downward trend, which suggests that the model is learning relevant patterns instead of overfitting. The learning rate remained constant at 1e-4 throughout, which supported smooth and steady training without the need for adjustments. Each epoch took approximately 18–20 seconds, pointing to stable processing times. Accuracy improvements across epochs were consistent and significant in both training and validation datasets. There were no noticeable signs of overfitting or

instability, and the model continued to learn efficiently across all epochs. These observations suggest that the model has reached a strong level of performance. The training process appears well-optimized and effective. The final evaluation of both training sets indicates the model is performing reliably. It is ready for deployment or advanced fine-tuning. The results affirm that the model is capable of generalizing well, with excellent predictive capabilities. With improved preprocessing methods, the dual-model framework supports real-time applications and attains high diagnostic precision. The model's potential as an additional diagnostic tool in clinical settings is confirmed by its balanced class-wise performance metrics and robustness across multiple datasets.

Future Work

This model can be used in real-time clinical settings in the future to facilitate quick medical diagnosis and decision-making. Integrating the framework into mobile diagnostic apps for the iOS and Android platforms is one important avenue that will allow medical practitioners to conduct real-time Chest X-ray analysis in emergency situations or remote locations. Furthermore, the model's lightweight design makes it perfect for implementing edge-AI on gadgets like the Raspberry Pi, NVIDIA Jetson Nano, and Google Coral. These gadgets can be integrated into mobile health vans or portable diagnostic kits that operate in remote or underdeveloped areas. To help radiologists with automated pre-screening and flagging of abnormal cases in emergency rooms, the system may also be integrated into hospital PACS (Picture Archiving and Communication Systems) and HIS (Hospital Information Systems). Future developments of this work will concentrate on integrating explainable AI tools like Grad-CAM++ to enhance model interpretability and clinical trust, extending the framework to other imaging modalities like CT and MRI, and fusing radiographic analysis with clinical metadata for multi-modal diagnosis. Additionally, federated learning might be used in subsequent studies to train the model across several hospital networks without jeopardizing data privacy. To improve model adaptability in low-data or rare-disease scenarios, developments in self-supervised and few-shot learning could also be investigated. Last but not least, integrating the model with wearable technology and Internet of Things-based clinical systems may facilitate ongoing observation and AI-assisted care delivery in outpatient and home environments.

Acknowledgement

The authors sincerely thank the Department of Information Technology at Annamalai University for their unwavering support, direction, and encouragement during this research project. In addition, the authors would like to thank the COVID-CXR Image Dataset (Research), which was essential to the training and validation of the suggested deep learning models. Finally, the authors express their gratitude to the reviewers for their insightful comments, which greatly enhanced the overall caliber and scientific rigor of this work.

Funding Information

The authors affirm that no funding from institutional, commercial, or governmental sources was used in the independent execution of this work. This study, data collection, model development, and manuscript preparation were all done without any funding.

Author's Contributions

Sriramkumar R: Led the conceptualization and execution of the research work. Developed the dual-model deep learning framework (MobileNetV2 + InceptionV3), conducted image preprocessing (CLAHE, white balance), performed dataset preparation, training, and validation. Authored the first draft and handled result interpretation and manuscript revisions.

K. Selvakumar: Provided expert supervision, critical insights, and research guidance throughout the study. Reviewed the methodology design, ensured the scientific integrity of the experimental workflow, and contributed to manuscript refinement and technical validation.

J. Jegan: Supported the analytical model evaluation, helped in the selection of appropriate CNN architectures, and contributed to literature review and editing. Assisted in data interpretation and formatting for journal submission.

Ethics

Since this study used pre-existing, anonymized, publicly available datasets, ethical approval was not necessary. There were no direct human or animal involvements. The authors attest that every piece of data was handled sensibly and in accordance with ethical research and data privacy guidelines.

References

Asghar, M. Z., Albogamy, F. R., Al-Rakhami, M. S., Asghar, J., Rahmat, M. K., Alam, M. M., Lajis, A., & Nasir, H. M. (2022). Facial Mask Detection Using Depthwise Separable Convolutional Neural Network Model During COVID-19 Pandemic. *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.855254

Boyina, K., Reddy, G. M., Akshita, G., Singh, R. P., Kumar, P., & Duraisamy, P. (2024). Detection and Analysis of Tuberculosis Disease from Chest X-Ray (CXR) Images using Machine Learning and Deep Learning Techniques. *2024 3rd International Conference for Innovation in Technology (INOCON)*, Bangalore, India. https://doi.org/10.1109/inocon60754.2024.10511723

Gaur, L., Bhatia, U., Jhanjhi, N. Z., Muhammad, G., & Masud, M. (2023). Medical image-based detection of COVID-19 using Deep Convolution Neural Networks. *Multimedia Systems*, *29*(3), 1729–1738. https://doi.org/10.1007/s00530-021-00794-6

Gunraj, H., Wang, L., & Wong, A. (2020). COVIDNet-CT: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases From Chest CT Images. *Frontiers in Medicine*, *7*. https://doi.org/10.3389/fmed.2020.608525

Hole, S. R., Bhavekar, G. S., Prajapati, A. K., Kolluru, V., Karpe, S. R., & Vrindavanam, J. (2025). Hybrid PCA-Based Machine Learning Models for Predictive Analytics in Urban Health Monitoring Systems. *2025 IEEE 1st International Conference on Smart and Sustainable Developments in Electrical Engineering (SSDEE)* https://doi.org/10.1109/ssdee64538.2025.10967575

Hole, S. R., Kolluru, V., Salotagi, S., Challagundla, Y., Mungara, S., & Sriramkumar, R. (2025). A Design of Hybrid Model and Bayesian Neural Networks for Smart Grid Stability Prediction. *2025 IEEE 1st International Conference on Smart and Sustainable Developments in Electrical Engineering (SSDEE)* https://doi.org/10.1109/ssdee64538.2025.10967715

Hole, S. R., Vrindavanam, J., Bhavekar, G. S., Naik Mude, N., Sriramkumar, R., Kolluru, V., & Hole, K. R. (2024). Impact of artificial intelligence on the development of employment and the labor market. *2024 International Conference on Augmented Reality, Intelligent Systems, and Industrial Automation (ARIIA)*, Manipal, India. https://doi.org/10.1109/ariia63345.2024.11051827

Huang, J., Mei, L., Long, M., Liu, Y., Sun, W., Li, X., Shen, H., Zhou, F., Ruan, X., Wang, D., Wang, S., Hu, T., & Lei, C. (2022). BM-Net: CNN-Based MobileNet-V3 and Bilinear Structure for Breast Cancer Detection in Whole Slide Images. *Bioengineering*, *9*(6), 261. https://doi.org/10.3390/bioengineering9060261

Hussain, A., Basak, R., & Mandal, S. (2024). *Deep Learning-Based Multiple Detection Techniques of Covid-19 Disease From Chest X-Ray Images Using Advanced Image Processing Methods and Transfer Learning*. 65–82. https://doi.org/10.1007/978-3-031-50815-8_5

Ifty, T. T., Shafin, S. A., Shahriar, S. M., & Towhid, T. (2024). Explainable Lung Disease Classification from Chest X-Ray Images Utilizing Deep Learning and XAI. *2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI)*, Mt Pleasant, MI, USA. https://doi.org/10.1109/icmi60790.2024.10586202

Ippolito, D., Maino, C., Gandola, D., Franco, P. N., Miron, R., Barbu, V., Bologna, M., Corso, R., & Breaban, M. E. (2023). Artificial Intelligence Applied to Chest X-ray: A Reliable Tool to Assess the Differential Diagnosis of Lung Pneumonia in the Emergency Department. *Diseases*, *11*(4), 171. https://doi.org/10.3390/diseases11040171

Krishnamoorthy, M., Muthusamy, S., Balakrishnan, B., Sridevi, S., Panneerselvam, M., Rameshkumar, M., & Salama, D. (2024). A novel multi class disease detection of chest x-ray images using deep learning with pre trained transfer learning models for medical imaging applications. Research Square. https://doi.org/10.21203/rs.3.rs-3946892/v1

Kumar, A., Nelson, L., & Gomathi, S. (2023). Sequential Transfer Learning Model for Pneumonia Detection Using Chest X-ray Images. *2023 Global Conference on Information Technologies and Communications (GCITC)*, Bangalore, India. https://doi.org/10.1109/gcitc60406.2023.10426367

Kumar Singh, V., Abdel-Nasser, M., Pandey, N., & Puig, D. (2021). LungINFseg: Segmenting COVID-19 Infected Regions in Lung CT Images Based on a Receptive-Field-Aware Deep Learning Framework. *Diagnostics*, *11*(2), 158. https://doi.org/10.3390/diagnostics11020158

Lakshmanan, M., Dhanraj, J. A., Sriramkumar, R., Naik, M. N., Mithaguru, & Godhandaraman, T. (2025). Blockchain-enabled medical waste managem Blockchain for Secure and Efficient Crowdfunding: An Optimized Particle Swarm Approachent system for enhanced traceability, safety and environmental protection,"Int. *J. Adv. Soft Comput. Appl*, *17*(2), 247–269. https://doi.org/10.15849/IJASCA.250730.13

Lakshmanan, M., Justindhas, Y., Ahamed, A. S., & Moorthy, A. (2022). Survey on Autonomous Vehicles using Artificial Intelligence. *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India. https://doi.org/10.1109/icesc54411.2022.9885422

Lakshmanan, M., Mala, G. S. A., Sivakumar, G., Divya Sundar, V. S., Kesavan, R., & Ajoe Sweetlin Jeena, A. (2024). A Secure and Efficient Framework for Storing Medical Data using Blockchain and Cloud Servers. *2024 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India. https://doi.org/10.1109/iceca63461.2024.10800855

Lakshmanan, M., Sriramkumar, R., Justindhas, Y., Kumar, H. . P., Mithaguru, & Ilamurugan, G. (2025). Blockchain-Based HSFO Framework for Privacy Preservation of Health Care Data Using Hybrid Algorithms. *2025 2nd International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*, Chennai, India. https://doi.org/10.1109/rmkmate64874.2025.11042708

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Lakshmanan, M., Mala, G. S. A., Poorni, R., Ilamurugan, G., Sriramkumar, R., & Gnanavel, R. (2024). Blockchain for Secure and Efficient Crowdfunding: An Optimized Particle Swarm Approach. *2024 9th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India. https://doi.org/10.1109/icces63552.2024.10860104

Maheswari, B. U., Sam, D., Mittal, N., Sharma, A., Kaur, S., Askar, S. S., & Abouhawwash, M. (2024). Explainable deep-neural-network supported scheme for tuberculosis detection from chest radiographs. *BMC Medical Imaging*, *24*(1). https://doi.org/10.1186/s12880-024-01202-x

Maquen-Niño, G. L. E., Nuñez-Fernandez, J. G., Taquila-Calderon, F. Y., Adrianzén-Olano, I., De-La-Cruz-VdV, P., & Carrión-Barco, G. (2024). Classification Model Using Transfer Learning for the Detection of Pneumonia in Chest X-Ray Images. *International Journal of Online and Biomedical Engineering (IJOE)*, *20*(05), 150–161. https://doi.org/10.3991/ijoe.v20i05.45277

Mazhar, T., Haq, I., Ditta, A., Mohsan, S. A. H., Rehman, F., Zafar, I., Gansau, J. A., & Goh, L. P. W. (2023). The Role of Machine Learning and Deep Learning Approaches for the Detection of Skin Cancer. *Healthcare*, *11*(3), 415. https://doi.org/10.3390/healthcare11030415

Naitik, S. T., & Gorabal, J. V. (2024). Design and Development of Multimodal Biometric System Using Finger Veins and Iris by CNN Integrated with Hybrid SIO and Whale Optimization Techniques. *International Journal of Interactive Mobile Technologies (IJIM)*, *18*(22), 97–114. https://doi.org/10.3991/ijim.v18i22.50865

Ouerhani, A., Boulares, S., & Mahjoubi, H. (2023). Automated Detection of Pediatric Pneumonia from Chest X-Ray Images Using Deep Learning Models. *2023 IEEE Afro-Mediterranean Conference on Artificial Intelligence (AMCAI)* https://doi.org/10.1109/amcai59331.2023.10431520

Saritha, S., Srinivas, V. S., Anuhya, D., & Pavithra, G. Performance Analysis of Detection of Disease on Leaf Images with Inception V3 and Mobilenet Deep Learning Techniques. (2022). *Journal of Pharmaceutical Negative Results*, *13*(SO3). https://doi.org/10.47750/pnr.2022.13.s03.018

Siddhartha, M. & Santra, A. (2020). COVIDLite: A depth-wise separable deep neural network with white balance and CLAHE for detection of COVID-19. *ArXiv:2006.13873* .

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., & Ng, A. Y. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *ArXiv:1711.05225*.

Sagar, A., Kolluru, V., Jaiswal, U., Kumavat, G., Hole, S. R., & Kumar, A. (2025). Machine Learning and Artificial Intelligence for Predictive Modeling in Antimicrobial Resistance Data Sets, Challenges, and Future Directions. *2025 3rd International Conference on Smart Systems for Applications in Electrical Sciences (ICSSES)*, Tumakuru, India. https://doi.org/10.1109/icsses64899.2025.11009316

Saood, A., & Hatem, I. (2021). COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet. *BMC Medical Imaging*, *21*(1). https://doi.org/10.1186/s12880-020-00529-5

Seekarajapuram Dinakaran, N., Muthusamy, L., & Vellore Sundaresan, D. S. (2025). Implementation of Energy-Aware Optimal Routing for Improving Traffic Capacity in Ad Hoc Wireless Network Using Hybrid Heuristic Algorithm. *International Journal of Communication Systems*, *38*(4). https://doi.org/10.1002/dac.6126

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, *35*(5), 1285–1298. https://doi.org/10.1109/tmi.2016.2528162

Singh, V. P., & Srivastava, M. K. (2024). Efficient COVID-19 Detection From Chest X-Ray Images Using Deep Transfer Learning. *Educational Administration Theory and Practices*. https://doi.org/10.53555/kuey.v30i5.4691

Sriramkumar, R., Selvakumar, K., & Jegan, J. (2025). Advances in Ai for Pulmonary Disease Diagnosis Using Lung X-ray Scan and Chest Multi-slice Ct Scan. *Journal of Theoretical and Applied Information Technology*, *103*(7).

Teixeira, L. O., Pereira, R. M., Bertolini, D., Oliveira, L. S., Nanni, L., Cavalcanti, G. D. C., & Costa, Y. M. G. (2021). Impact of Lung Segmentation on the Diagnosis and Explanation of COVID-19 in Chest X-ray Images. *Sensors*, *21*(21), 7116. https://doi.org/10.3390/s21217116

Thimoteo, L. M., Vellasco, M. M., Amaral, J., Figueiredo, K., Yokoyama, C. L., & Marques, E. (2022). Explainable Artificial Intelligence for COVID-19 Diagnosis Through Blood Test Variables. *Journal of Control, Automation and Electrical Systems*, *33*(2), 625–644. https://doi.org/10.1007/s40313-021-00858-y

World Health Organization (2022). Global Tuberculosis Report 2021: Supplementary Material.

Zanaj, E., Balliu, L., Basha, G., Gjata, E., & Meçe, E. K. (2024). Studying the Behavior of a Modified Deep Learning Model for Disease Detection Through X-ray Chest Images. *International Journal of Advanced Computer Science and Applications*, *15*(5). https://doi.org/10.14569/ijacsa.2024.0150585